TITLE	**Framework for Processing Videos in the Presence of Spatially Varying Motion Blur**
PI	Prof. A.N. Rajagopalan, Indian Institute of Technology Madras
AFRL POC	Dr. Guna Seetharaman, Civ DR-IV AFRL/RIEA
AOARD PM	Dr. Seng Hong
AFOSR PM	Dr. Tristan Nguyen, AFOSR/RSL
Duration	Sept. 1, 2012 - August 31, 2013
Cost	50K (FY13)

1 Introduction

This report is focused on a basic (6.1) level research on full motion video analysis with the objective of investigating the process of motion blurring in planar scenes and to develop algorithms for blind restoration of blurred images together with an estimate of camera motion. This is a topic of immense importance to the U.S. Air Force - with a potential impact on image analysis, characterization and exploitation. The amount of full motion video clips that we process has grown exponentially. These images are typically acquired for surveillance purpose, collected persistently over a fixed field of view, albeit with varying degrees of relative motion between the camera and objects within the scene. Motion blur results when there is relative motion between camera and scene. It has acquired special significance with hand-held imaging, aerial imaging, and imaging 'on the move' shooting into prominence. It is also relevant to situations where the camera is still but the scene comprises of several moving objects. The problem is of great relevance within the overall context of aerial surveillance. Images captured from an aerial platform are normally affected by motion blur due to instabilities of the moving platform. Smear caused by motion blur can severely dent the utility of such data.

The ubiquity of imaging devices and the fact that practical cameras are real-aperture results in the natural prevalence of motion/optical blur in images [1]. The problem of image blurring has been around for many years, the major

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
18 APR 2014	**Final**	**20-08-2012 to 19-08-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Framework for Processing Videos in the Presence of Spatially Varying Motion Blur	**FA23861214070**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Ambasamudram Narayanan Rajagopalan	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Indian Institute of Technology Madras,IIT Madras,Chennai 600036,India,IN,600036	**N/A**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
AOARD, UNIT 45002, APO, AP, 96338-5002	**AOARD**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	AOARD-124070

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The research is continued for two more years.

14. ABSTRACT
The problem of image blurring has been around for many years, the major thrust having been on optical defocusing that occurs when the lens settings relative to a scene fail to satisfy Gauss law. Depending on the nature of the scene, the resultant image can be either space-invariantly or space-variantly blurred. Several works already exist that deal with optical blur in a comprehensive manner.

15. SUBJECT TERMS
full motion video analysis, Image Processing, Video analysis, Information Technology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**38**	

thrust having been on optical defocusing that occurs when the lens settings relative to a scene fail to satisfy Gauss law. Depending on the nature of the scene, the resultant image can be either space-invariantly or space-variantly blurred. Several works already exist that deal with optical blur in a comprehensive manner. Interestingly, recent times have seen the resurgence of motion blur as an area of great relevance. In contrast to defocus, the occurrence of motion blur is significantly higher in practical scenarios. Motion blur can be avoided provided the camera is placed on a firm support. However, carrying sturdy accessories can be cumbersome and sometimes prohibitive. Even if one were to use auxiliary sensors, blur cannot be completely avoided [2].

Unlike the optical blur, motion-blur can be space-varying even for planar scenes and is typically so since camera motion invariably involves rotations. An example case is that of a rotating camera imaging a distant scene. In fact, the shape of the blur kernel is a function of the scene as well as camera motion. For planar scenes, the shape of the blur kernel is a function of camera motion while the weights of the kernel can be related to the exposure time corresponding to the set of geometric transformations that the camera traversed along its motion trajectory. Unlike the optical blur, the motion blur is a function of both camera motion and scene depth, has no a priori shape, and is typically sparse. These characteristics make the motion blur unique in several respects as compared to the traditionally well-understood optical blur.

Motion blurring is both a bane and a boon. Most works treat motion blur as nuisance and seek ways and means to mitigate its effects so as to restore the original image. However, it must be emphasized that motion blur can also serve as a vital cue for camera motion estimation, depth recovery, super-resolution, image forensics, etc. Blind motion deblurring is both interesting and a technically challenging problem. The aim of this project was to study, understand, investigate, and propose algorithms to meet our objectives primarily from a surveillance (not necessarily aerial) perspective. The key challenge lay in handling the complexities (including loss of resolution) that arise from space-varying local blurring due to incidental camera motion.

We started joint work with the AFRL team led by Dr. Guna Seetharaman in Sept. 2012. Our focus was on restoration, registration, dehazing, and super-resolution of motion blurred images for planar scenes. There were frequent information exchanges between the P.I. and the AFRL collaborator, at least once every quarter. The analytical work focused on arriving at global explanations for the blurring process by modeling the motion-blurred image as an average of projectively transformed images. Initially, we address restoration of space-variantly motion-blurred images. This was followed by a method for registering blurred images to reveal change detection, if any. We then discuss how to restore foggy motion-blurred images using depth cues derived from fog itself. Finally, for purpose

of image super-resolution, the deblurring framework was extended to handle down-sampling effects too. For each of these tasks, once the analytical formulations were in place, the simulation cum validation phase followed with tests involving real data sets. Experimental verification of the algorithms, including computer simulations, was done to refine both the theoretical analysis as well as the numerical implementation. While some of the results arising out of these efforts have already been published in prestigious avenues, others are under review.

2 Restoration of motion blur

During the process of capturing an image, the camera shutter opens and closes to let light from the scene to fall on imaging sensors. The shutter interval denotes the amount of time during which the camera sensors observe the scene. The final image obtained from the camera is a function of the total light energy accumulated by the sensors. When the shutter is open, if there is a motion of either the camera or the scene, a particular pixel in the image plane receives light intensities from more than one point in the scene resulting in an averaging effect called as motion blur [1, 3, 4].

Let f denote the image captured from a camera when there is no relative motion during exposure (shutter interval). i.e., f represents the unblurred latent image of the scene. Let g denote a blurred image captured when there is a relative motion between the camera and the scene. The intensity value at a pixel of g will be an average of image intensities at different pixels of f i.e.,

$$g(x,y) = \frac{1}{T_e} \int_0^{T_e} f\left(x - \overline{x}(x,y,\tau), y - \overline{y}(x,y,\tau)\right) \, d\tau \tag{1}$$

where T_e is the exposure time, and $\overline{x}(x,y,\tau)$ and $\overline{y}(x,y,\tau)$ denote the components of the apparent displacement of a point (x,y) in f at time instant τ. In the special case where the displacement is the same for all the image points, we have $\overline{x}(x,y,\tau) = x_0(\tau)$, $\overline{y}(x,y,\tau) = y_0(\tau)$ and

$$g(x,y) = \frac{1}{T_e} \int_0^{T_e} f\left(x - x_0(\tau), y - y_0(\tau)\right) \, d\tau \tag{2}$$

All the image points undergo the same displacement only when the relative motion is restricted to in-plane translations and when there is no effect of parallax (i.e., scene is a fronto-parallel plane). In this scenario, motion blur can be modeled as a convolution of the original image with a point spread function (PSF) which is also referred to as a blur kernel [3, 5]. i.e.,

$$g(x,y) = f * h(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - s, y - t) \, h(s,t) \, ds \, dt \tag{3}$$
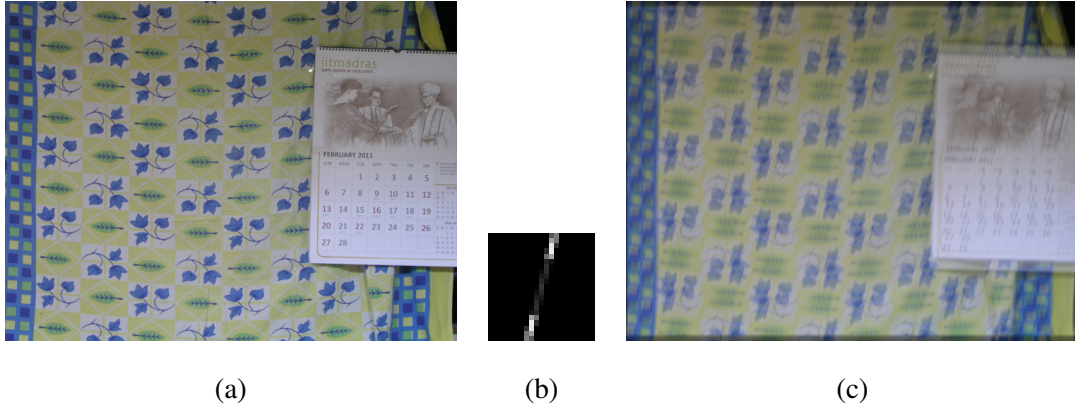
3

Figure 1: (a) Unblurred latent image of the scene. (b) Blur kernel. (c) Blurred observation.

The PSF h is given by

$$h(s,t) = \frac{1}{T_e} \int_0^{T_e} \delta\left(s - x_0(\tau), t - y_0(\tau)\right) \, d\tau \tag{4}$$

where δ indicates the 2D Dirac-delta function. An interesting interpretation of equation (3) is that the blurred image g is a weighted average of translated instances of the original image f. The PSF h denotes the weights corresponding to every possible translational shift of the camera. The value of the weight at a position is equal to the fraction of the exposure duration spent by the camera in that shifted position. In the special case where the camera velocity is uniform and along a particular direction, the PSF will have uniform weights only along a straight line whose angle denotes the direction of motion. Note that the PSF cannot be negative. Since the relative motion is arbitrary, we cannot represent a motion blur kernel in terms of a predetermined functional form. Also, motion PSF is typically sparse because the displacements undergone by a scene-point form only a small subset of all possible shifts. If the exposure duration is the same for the original image as well as the blurred image, the magnitude of light energy accumulated will be the same for both the observations. Hence, by principle of energy conservation, the PSF integrates to unity.

In Fig. 1, we show the effect of space-invariant motion blurring. The original image of a scene was captured from a still camera as shown in Fig. 1 (a). In Fig. 1 (c), we show the motion blurred image obtained when there was a camera shake during exposure. The blur kernel given in Fig. 1 (b) depicts the direction of camera translation. Since the camera motion was predominantly along vertical direction, we observe smearing effect along that direction in Fig. 1 (c).

2.1 Motion blur and TSF model

When the camera motion is not restricted to in-plane translations, the apparent motion of scene points in the image will vary at different locations resulting in space-variant blurring. The convolution model with a single blur kernel does not hold in such a scenario. However, when the scene depth is constant, the blurred image can be accurately modeled as a weighted average of the warped instances of the original image [6, 7, 8, 9].

Let the image of a scene captured by a still camera be denoted by $f : \mathbb{R}^2 \to \mathbb{R}$. Let $\mathbf{X} = [X \ Y \ Z]^T$ denote the spatial coordinates of a point in the scene with the camera center as the origin. The projection of $\mathbf{X}$ in the image plane (x, y) is given by $x = \frac{\nu X}{Z}$ and $y = \frac{\nu Y}{Z}$ where ν denotes the focal length. Using homogeneous coordinates, the image point $\mathbf{x} = [x \ y \ 1]^T$ can be written as $K_\nu \mathbf{X}$. In this discussion, K_ν is assumed to be of the form

$$K_\nu = \begin{bmatrix} \nu & 0 & 0 \\ 0 & \nu & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

Due to camera motion during image capture, at each instant of time τ, the coordinates of the 3D point $\mathbf{X}$ changes to $\mathbf{X}_\tau = R_\tau \mathbf{X} + T_\tau$ with respect to the camera where $T_\tau = [T_{X_\tau} \ T_{Y_\tau} \ T_{Z_\tau}]^T$ is the translation vector. The rotation matrix R_τ is parameterized [10] in terms of θ_X, θ_Y and θ_Z (the angles of rotation about the three axes) using the matrix exponential

$$R_\tau = e^{\Theta_\tau} \text{ where } \Theta_\tau = \begin{bmatrix} 0 & -\theta_{Z_\tau} & \theta_{Y_\tau} \\ \theta_{Z_\tau} & 0 & -\theta_{X_\tau} \\ -\theta_{Y_\tau} & \theta_{X_\tau} & 0 \end{bmatrix} \tag{6}$$

We consider that all of the scene points are at a depth d_o from the camera. Consequently, the point $\mathbf{x}_\tau$ at which $\mathbf{X}_\tau$ gets projected in the camera can be obtained through a homography H_τ as $\mathbf{x}_\tau = H_\tau \mathbf{x}$ where

$$H_\tau = K_\nu \left(R_\tau + \frac{1}{d_o} T_\tau [0 \ 0 \ 1] \right) K_\nu^{-1} \tag{7}$$

Let g_τ denote the image captured at time instant τ. For the sake of simplicity, we use the same notation ($\mathbf{x}$) for the homogeneous coordinates as well as for the coordinates in the image plane. Then we can write $g_\tau(\mathbf{x}) = f(H_\tau^{-1} \mathbf{x})$ where H_τ^{-1} denotes the inverse of H_τ (since $g_\tau(H_\tau \mathbf{x}) = f(\mathbf{x})$). The blurred image g can be considered as the average of the light intensities observed in the image plane during exposure. The blurred image intensity at an image

point $\mathbf{x}$ is given by

$$g\left(\mathbf{x}\right) = \frac{1}{T_e} \int\limits_0^{T_e} f\left(H_\tau^{-1}\mathbf{x}\right) d\tau \tag{8}$$

where T_e is the total exposure duration.

Note that, when averaging over time, the temporal information (order of the set of transformations undergone by the reference image) is lost. But this is a non-issue for the problem on hand. The blurred image can be more appropriately modeled in terms of the reference image, by averaging it over the set of possible transformations (resulting from the camera motion). Let $\mathcal{T}$ denote the set of all possible transformations and $\mathbf{\Upsilon}$ denote a transformation. We define the *transformation spread function* (TSF) $\omega : \mathcal{T} \to \mathbb{R}_+$ as a mapping from the set $\mathcal{T}$ to non-negative real numbers. For each transformation $\mathbf{\Upsilon} \in \mathcal{T}$, the value of the TSF $\omega\left(\mathbf{\Upsilon}\right)$ denotes the fraction of the total exposure duration for which the camera was in the position that caused the homography $H_{\mathbf{\Upsilon}}^{-1}$ on the image coordinates. It is to be noted that the term $\mathbf{\Upsilon}$ denotes the transformation parameters corresponding to the homography matrix $H_{\mathbf{\Upsilon}}^{-1}$, and does not indicate a time instant. The blurred image can be written as an average of the warped images weighted by the TSF ω. i.e.,

$$g\left(\mathbf{x}\right) = \int\limits_{\mathbf{\Upsilon} \in \mathcal{T}} \omega\left(\mathbf{\Upsilon}\right) f\left(H_{\mathbf{\Upsilon}}^{-1}\left(\mathbf{x}\right)\right) d\mathbf{\Upsilon} \tag{9}$$

When the camera motion is not restricted, the paths traced by scene points in the image plane can vary across the image resulting in space-variant blur. However, the blurring operation can be described by a single TSF using equation (9). The TSF depicts the camera motion during exposure. For instance, if the camera undergoes only in-plane rotations, the TSF will have non-zero weights only for the rotational transformations. Analogous to a blur kernel, the TSF satisfies the relation $\int\limits_{\mathbf{\Upsilon} \in \mathcal{T}} \omega\left(\mathbf{\Upsilon}\right) = 1$ (assuming equal amount of light energy is involved in the formation of f and g).

Although, the TSF model is more efficient than the PSF model in representing space-variant motion blur, it is useful to relate the two models. The blurred image g can be modeled with a space-variant PSF h as

$$g\left(\mathbf{x}\right) = f *_v h\left(\mathbf{x}\right) = \int f\left(\mathbf{x} - \mathbf{u}\right) h\left(\mathbf{x} - \mathbf{u}, \mathbf{u}\right) d\mathbf{u} \tag{10}$$

where $h\left(\mathbf{x}, \mathbf{u}\right)$ denotes the blur kernel at the image point $\mathbf{x}$ as a function of the independent variable $\mathbf{u}$. The PSF $h\left(\mathbf{x}, \mathbf{u}\right)$ represents the displacements undergone by a point light source at $\mathbf{x}$ during the exposure and is weighted according to the fraction of the exposure time the light source stays at the displaced position. Following our discussions

in section 2, the PSF can be written as

$$h\left(\mathbf{x}, \mathbf{u}\right) = \frac{1}{T_e} \int\limits_0^{T_e} \delta\left(\mathbf{u} - \bar{\mathbf{x}}_\tau\right) d\tau \tag{11}$$

where δ indicates the 2D Dirac Delta function and $\bar{\mathbf{x}}_\tau$ is the instantaneous displacement. It is to be noted that for brevity, we use the vectors $\mathbf{x}$ and $\mathbf{u}$ to indicate locations. The PSF h at an image point $\mathbf{x}$ can be obtained from the TSF ω by finding the displacement induced due to each of the possible transformations. This relationship can be written as

$$h\left(\mathbf{x}, \mathbf{u}\right) = \int\limits_{\mathbf{\Upsilon} \in \mathcal{T}} \omega\left(\mathbf{\Upsilon}\right) \delta\left(\mathbf{u} - \left(H_{\mathbf{\Upsilon}}\mathbf{x} - \mathbf{x}\right)\right) d\mathbf{\Upsilon} \tag{12}$$

Consider the scene whose latent image is shown in Fig. 2 (a). There were no depth variations in this scene. However, the camera underwent in-plane rotational motion during image capture leading to the non-uniformly blurred image shown in Fig. 2 (b). We note that the region around the image center does not appear blurred while the effect of blurring increases as we move away from the center.
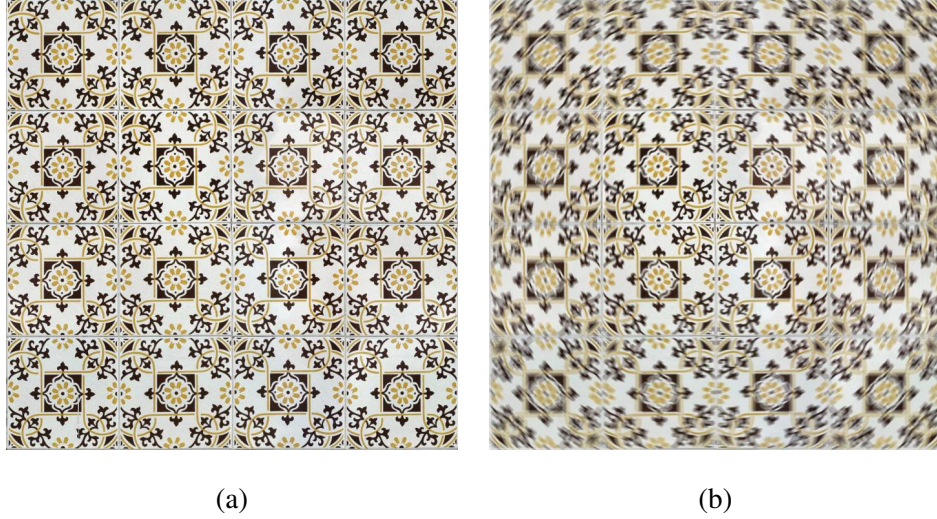

(a)   (b)

Figure 2: (a) Unblurred image. (b) Space-variantly blurred observation.

2.2 Related works

The problem of inferring an image in the presence of blur (which corresponds to image deblurring) has been widely studied in the literature [4, 7, 11, 12, 13]. Traditional approaches assume blur to remain constant across all the image points. This will be the case when the camera motion is restricted to in-plane translations and the scene is of constant

depth. A lot of approaches that address the problem of blind motion deblurring exist in the literature for the case of uniform blur [4, 14, 15, 16]. A comparison of recent deconvolution techniques for space-invariant blur can be found in [13]. In practical scenarios, blurring due to camera shake is space-variant in nature and the convolution model does not hold [17]. [18] have proposed a non-blind deblurring scheme which estimates the latent image even if the blur kernels are erroneous and demonstrated its applicability for space-variant blur. Based on this work, in [19], they address blind space-variant deblurring by estimating the blur kernel at each pixel. Techniques exist that address restoration of non-uniform blur by local space-invariance approximation [20, 21]. Recent techniques avoid such an approximation by modeling the motion-blurred image as an average of projectively transformed images [6, 7, 8, 9]. In this approach, blur is modeled by considering the transformations undergone by the image plane rather than using a point spread function which varies at every pixel. [6] have proposed a deblurring scheme for the projective blur model based on Richardson Lucy deconvolution. However, they do not address the problem of determining the blurring function. [7, 10] propose an image restoration technique for non-uniform motion-blur arising due to camera rotations. They represent the blurring function on a 3D grid corresponding to the three directions of camera rotations. For the case of blind image restoration, the kernel estimation framework is employed in [4]. When a noisy version of the original image is available, a least-squares energy minimization approach is used for finding the blurring function. [9] have proposed another blind deblurring scheme wherein the camera motion is considered to be comprised of 2D translations and in-plane rotation. Their technique is based on the method in [14]. In the restoration technique by [22], sensors are attached to the camera to determine the blurring function. [23] propose a deblurring scheme that uses coded exposure and some simple user interactions to determine the space-variant PSF. [24] propose a new approach to restore non-uniform motion blur by using the efficient filter flow framework [25]. In [26], a new regularization scheme that compensates attenuation of high frequencies is used to perform blind deblurring. While [27] uses two non-uniformly blurred observations and iteratively solve for the latent image and its transformations, in [28], the authors propose to iteratively estimate the latent image and the blur from a single image.

2.3 Blind image restoration

Our algorithm proceeds by updating the estimate of the camera pose at one step, and the latent image in the next. To this end, we minimize the following energy function. Let $\omega = h_T$ and $l = f$. The cost to be minimized is

$$E(\mathbf{w}, \mathbf{l}) = \left\|\left(\sum_{k \in \mathbf{T}} \omega(k) \cdot \mathbf{l}(\mathbf{H_k}x)\right) - \mathbf{b}\right\|_2^2 + \Phi_1(\mathbf{l}) + \Phi_2(\mathbf{w}) \tag{13}$$

where $\mathbf{w}$ denotes the vector of weights $\omega(k)$, Φ_1 and Φ_2 represent the regularization terms on the latent image $\mathbf{l}$ and the weights $\mathbf{w}$, respectively. The regularization terms will be explained in the following sub-sections.

2.3.1 TSF estimation

In the TSF estimation step, we compute $\mathbf{w}$ given the current estimate of the latent image $\mathbf{l}$ based on equation 13. Our algorithm requires the user to specify a rough guess of the extent of the blur (translation in pixels along x,y axes and rotation in degrees along z axis) to build the initial TSF. The three-dimensional camera pose subspace, whose limits are specified by the user, is uniformly sampled to build the first 'active' set of camera poses. We denote this active set by $\mathbf{A}$ where $\mathbf{A} \subset \mathbf{T}$. (In our experiments, the active set contained 200 poses which is still much smaller than the 1500-2000 poses that the whole space $\mathbf{T}$ would contain even for small to moderate blurs.) Our algorithm requires no other user input.
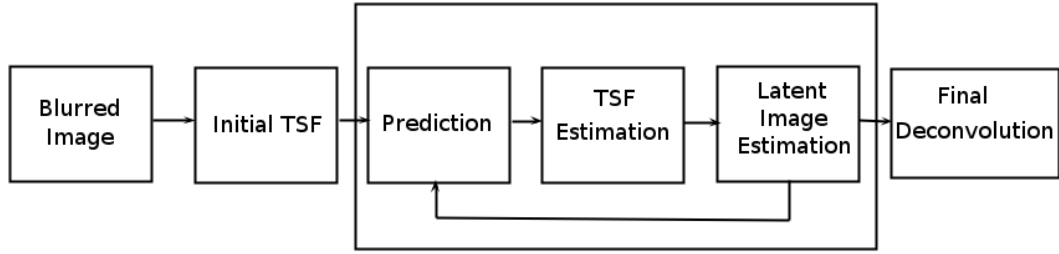

Figure 3: Overview of the proposed deblurring algorithm.

In the first iteration, we optimize over the initial TSF (obtained in 2.3.1) by minimizing the following energy function

$$E(\mathbf{w}) = \left\lVert \left(\sum_{k \in \mathbf{A}} \omega(k) \cdot \mathbf{l}(\widehat{\mathbf{H}_k} x) \right) - \mathbf{b} \right\rVert_2^2 + \Phi_2(\mathbf{w}) \tag{14}$$

where $\Phi_2(\mathbf{w}) = \beta \lVert \mathbf{w} \rVert_1$, a sparse prior. Since image derivatives have been shown to be effective for reducing ringing effects [14], we work on gradients instead of image intensities in our implementation of equation 14. This optimization problem can be solved using the nnLeastR function of the Lasso algorithm [29] which considers the additional $l1 - norm$ constraint and imposes non-negativity on the TSF weights. Only the 'dominant' poses in the active set $\mathbf{A}$ are selected as a result of the sparsity constraint imposed by the $l1 - norm$ and the remaining poses which are outliers are removed. We now rebuild the active set for the next iteration so that its cardinality is the same as the $\mathbf{A}$ of the current iteration. The new poses are selected, in a manner similar to [28], by sampling based on the

current $\mathbf{A}$ using a Gaussian distribution. In the next iteration, equation 14 is minimized over this new active set. The variance of the Gaussian distribution is gradually reduced with iterations as the estimated TSF converges to the true TSF. Experiments on synthetic and real data have shown that our algorithm exhibits good convergence properties and does not get stuck in local minima. We used a β value of 0.5 for our experiments.

2.3.2   Latent image estimation

We first perform image prediction at each iteration before the TSF estimation step to obtain more accurate results and to facilitate faster convergence. The prediction step consists of bilateral filtering, shock filtering and gradient magnitude thresholding. The predicted image, denoted by $\widehat{\mathbf{l}}$ is sharper than the estimated latent image $\mathbf{l}$ from the previous iteration and has fewer artifacts.

The latent image $\mathbf{l}$ is estimated by fixing the TSF weights $\mathbf{w}$. The blurring matrix is constructed using only the poses in the active set since the weights of the poses of the inactive set are zero, i.e. $\widehat{\mathbf{H}} = \sum_{k \in \mathbf{A}} \omega(k) \mathbf{H_k}$ and the energy function to be minimized takes the form

$$E(\mathbf{l}) = ||\widehat{\mathbf{H}}\mathbf{l} - \mathbf{b}||_2^2 + \Phi_1(\mathbf{l}) \tag{15}$$

We use the regularization terms in [14] and a conjugate gradient approach for optimization. An overview of our method is given in Fig. 3.

2.4   Results

We tested out method on synthetic and real data. Fig. 4 shows a synthetic case where a focused original image was blurred using a randomly generated 3D TSF to obtain the blurred image in Fig. 4(a). The parameters of the TSF ranged as follows: θ ranged between -1.5 to 1.5 degrees in steps of 0.25, t_x and t_y ranged between -4 to 4 pixels in steps of one pixel. The number of non-zero transformations in the TSF was set to 25. The poses in the TSF are defined in such a way that it depicts the path traversed by a camera with non-uniform velocity. A slightly overestimated blur size of -2 to 2 degrees along θ, -5 to 5 pixels along t_x and t_y was input to our deblurring algorithm. Our deblurred result (Fig. 4 (b)) is sharp and is free from artifacts.

The real examples shown in the first and second rows of Fig. 5 are aerial images from VIRAT dataset. The efficacy of our method in deblurring these real images is clearly evident from the artifact free and sharp output results.

<div align="center">(a)   (b)</div>

Figure 4: (a) Blurred observation. (b) Deblurred result.

3   Registration and occlusion detection in motion blur

Detecting occluded regions in images is an extensively studied problem in image processing and computer vision due to its applicability to a vast range of areas such as tracking, surveillance, object recognition, inpainting [30], [31], [32], [33], [34] etc. The objective, in a typical setting, is to automatically detect occlusion(s) given a pair of images taken from different view points and at different times. The occlusions themselves may have been caused by the entry or disappearance of objects in the scene within the time-span of the two observations. A common approach is to first compensate for the variations in pose by registering the two images with respect to each other followed by differencing to reveal changes in the scene. For small occlusions, the images can be aligned even using standard registration techniques [35], [36] that do not account for occlusions. This is because the matching of unoccluded pixels can be expected to sufficiently outweigh any possible degradation arising from attempting to match occluded pixels with unoccluded pixels and vice versa. However, larger occlusions warrant methods that detect the occluded pixels and exclude them from the registration process [37]. This challenging problem of detecting occlusions becomes even more ill-posed if one of the images in the pair is blurred due to the presence of camera shake. This is often the case when a quick fly-through is attempted for the recoverage of a particular geographic area, for which detailed surveillance images (i.e., latent images) are already available. Moreover, if the revisit is made at a time when the luminance is weak [38], then the exposure time needs to be increased, thereby increasing the chances of motion blur. Detecting occlusions is important for revealing changes in infrastructure, deployment of military units, modification/introduction of equipment etc. As pointed out in [39], traditional registration methods such as direct and feature-based approaches cannot be used in such a case due to photometric inconsistencies introduced by the blur. The alignment approach
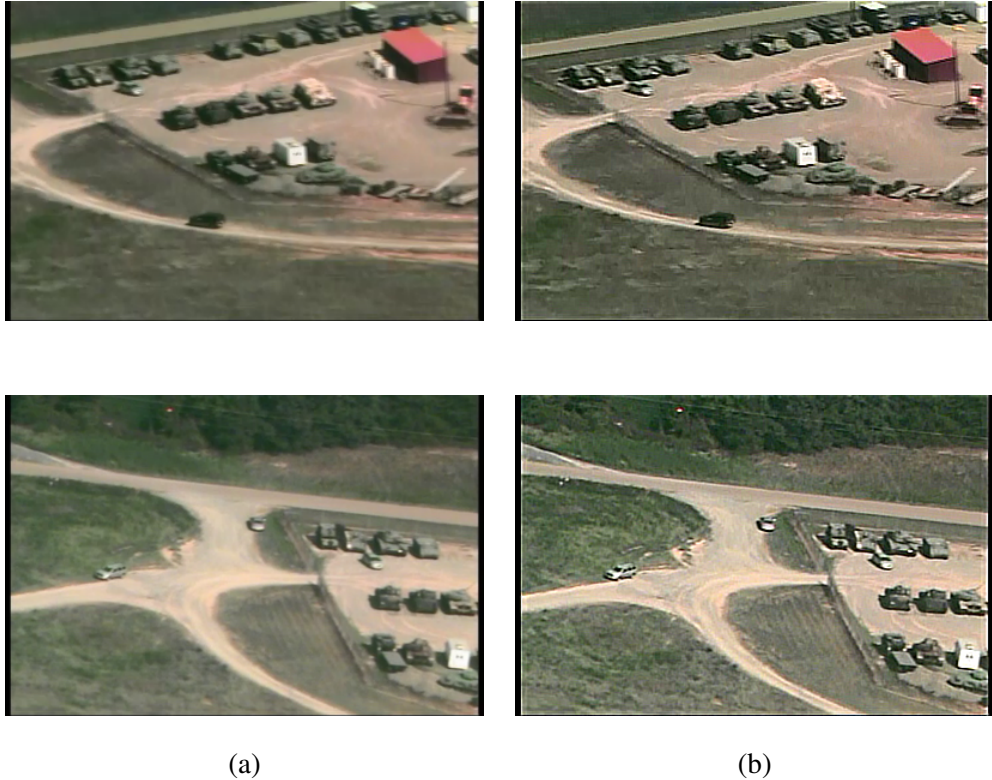
| (a) | (b) |

Figure 5: VIRAT dataset. (a) Motion-blurred real observations. (b) Deblurred results.

presented in [39] is based on the convolution model and applies only to the restrictive uniform blur case. However, in the case of general camera motion, the blur incurred can be significantly non-uniform across the image, and a space-varying formulation becomes necessary to describe the blurring process. It is this compounded scenario that we address in this work. Note that there can be more than one occluder.

While conventional approaches to detecting occluder(s) would require one to follow the deblur-register-difference pipeline, we present a unified framework which directly solves for the occluder(s) by accounting for the non-uniform blur and the changes in camera pose given a blurred/unblurred image pair. We show that direct registration of the pair is possible without the need for deblurring. Registration, in this context, tantamounts to estimating the set of warps which when applied on the focused image aligns it with the blurred image in the region of overlap. The elegance of our method lies in the fact that registration and occlusion detection turn out to be a natural fallout of our blur estimation process. We assume that the occluded pixels occupy only a relatively small portion of the image and that the camera motion trajectory is sparse in the camera motion space. We also assume the scene to be sufficiently far away so that depth variations can be ignored. We use a multiscale approach in which the image resolution is varied from

12

coarse-to-fine, thus rendering the algorithm efficient both in terms of computational time and memory requirements.

3.1   Sparsity, registration and occlusion handling

If $\mathbf{l}$, $\mathbf{b}$ represent the latent image and the blurred image, respectively, lexicographically ordered as vectors, then, in matrix-vector notation we can write

$$\mathbf{b} = A\boldsymbol{\omega} \tag{16}$$

where A is the matrix whose columns contain projectively transformed copies of $\mathbf{l}$, and $\boldsymbol{\omega}$ denotes the vector of weights $\omega(k)$. Note that $\boldsymbol{\omega}$ is a sparse vector since the blur is typically due to incidental camera shake and only a small fraction of the poses in $\mathbf{T}$ will have non-zero weights in $\boldsymbol{\omega}$.

In the scenario that we consider, one of the images in the pair is not only blurred because of camera jitter but can also contains occluder(s). To deal with this situation, we modify the linear model of (16) as

$$\mathbf{b_{occ}} = \mathbf{b} + \mathbf{o} \tag{17}$$

where $\mathbf{b_{occ}}$ is the blurred and occluded observation. In the image formation model, the occlusion happens first followed by blurring, i.e., $\mathbf{b_{occ}}$ is the weighted average of warped instances of an unknown focused image containing occlusions. The non-zero entries of $\mathbf{o}$, therefore, model the blurred occluder(s) in $\mathbf{b_{occ}}$. Since the occluder(s) can have arbitrary intensities, techniques designed for small noise cannot be used here. The locations of occlusion differ for different input images and are not known a priori to the algorithm. But we assume that the occluded pixels occupy only a relatively small portion of the image. Therefore, the occlusion vector $\mathbf{o}$, in the same vein as the vector $\boldsymbol{\omega}$, has sparse non-zero entries [40]. Since $\mathbf{b} = A\boldsymbol{\omega}$, we rewrite equation (17) as

$$\mathbf{b_{occ}} = \begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{o} \end{bmatrix} = B\mathbf{x} \tag{18}$$

Here $B = [A\ I] \in \mathbb{R}^{N \times (N_{\mathbf{T}} + N)}$, where N is the total number of pixels in the image, $N_{\mathbf{T}}$ is the total number of transformations in $\mathbf{T}$, and I is an $N \times N$ identity matrix. Hence, the system $\mathbf{b_{occ}} = B\mathbf{x}$ is always underdetermined and does not have a unique solution for $\mathbf{x}$. We, therefore, attempt to recover $\mathbf{x}$ as the sparsest solution to the system $\mathbf{b_{occ}} = B\mathbf{x}$. Note that in the absence of occlusion, $\mathbf{x} = \boldsymbol{\omega}$ and our problem reduces to the special case of estimating a sparse TSF.

With the occlusion model incorporated, the energy function to be minimized takes the form

$$E(\mathbf{x}) = ||\mathbf{b_{occ}} - B\mathbf{x}||_2^2 + \beta||\mathbf{x}||_1 \tag{19}$$

$$\text{s.t} \quad \forall k \in \mathbf{T}, \omega(k) \geq 0 \quad \text{and} \quad \sum_{k \in \mathbf{T}} \omega(k) = 1.$$

In the absence of an occluder, the convex combination of the elements of a particular row, say i, of A produces the intensity of the blurred pixel at the ith location in the image. If the observed intenstiy (in $\mathbf{b_{occ}}$) at the ith pixel is greater than the maximum intensity of the elements of the ith row, then, by convexity, we can deduce that it is the presence of a bright occluder that causes the intensity at that pixel to increase. A positive value in $\mathbf{o}$ will then explain the observed intensity at that pixel. On the other hand, if the observed intensity at the ith pixel is less than the minimum intensity of the elements of the ith row, we conclude that the occluder is dark. In this case, we replace the '1' at the corresponding location in I with a '-1'. This change in sign permits us to impose non-negativity on $\mathbf{x}$ because the residual can now take both positive and negative values. Thus B now becomes $[A \; I_{mod}]$ where I_{mod} is a diagonal matrix (with +1 and -1 along the diagonal) obtained after verifying the above condition.

3.2   Experiments

This section consists of two parts. We first evaluate the performance of our algorithm on synthetic data. Following this, we demonstrate the applicability of the proposed method on real images.

We begin with a synthetic example. A latent image of size 240×240 pixels of an airport bay is shown in Fig. 6(a). The same scene from a different camera pose and with synthetically added occluders (enclosed in red boxes) is shown in Fig. 6(b). The TSF space is chosen as follows- in-plane translations: T_x, $T_y = [-7 : 1 : 7]$, in-plane rotation: $R_z = [-3\circ : 1\circ : 3\circ]$, out-of-plane translation: $T_z = [0.95 : 0.05 : 1.05]$ and out-of-plane rotations: R_x, $R_y = [\frac{-4}{3}\circ : \frac{1}{3}\circ : \frac{4}{3}\circ]$. To simulate the motion of the camera, we manually generate 6D camera motion with a connected path in the motion space and initialize the weights. The synthezied camera motion (TSF model) is applied on Fig. 6(b) to produce the blurred and occluded image (Fig. 6(c)). To evaluate the proposed method, we set the number of scales in the multiscale implementation to 3 and first coarsely align the latent image and the blurred and occluded image at the lowest resolution without accounting for occlusion. In this step, the transformation intervals are expanded to T_x, $T_y = [-40 : 1 : 40]$ and $R_z = [-8\circ : 1\circ : 8\circ]$ to accommodate for the large change in pose between the two images. Note that this increase in the transformation intervals is not very demanding because we work at the lowest resolution of the image and the TSF in the multiscale algorithm. The 'dominant pose', i.e., the pose with the
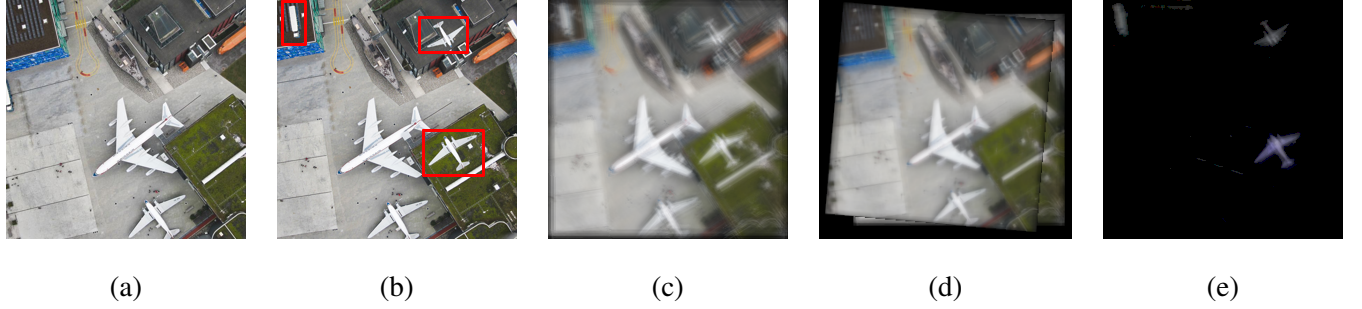
14

Figure 6: (a) Latent image, (b) latent image from a different camera pose and with synthetically added occluders, (c) blurred and occluded observation, (d) latent image reblurred using the estimated camera motion and overlaid on the blurred and occluded observation, and (e) residual image.

highest weight from the estimated vector ω is used to align the latent image with the blurred image. The TSF is now built around this dominant pose and we minimize equation (19) using the multiscale approach but now by also taking occlusions into consideration. Fig. 6(d) shows the latent image reblurred using the estimated ω and overlaid on the blurred and occluded observation. It is to be noted that the TSF model implicitly accounts for the change in pose between the two images. The residual image shown in Fig. 6(e) is the absolute difference between the blurred and occluded observation (Fig. 6(c)), and the reblurred latent image. Note that the occluders are correctly detected.

A real example but with an appreciable change in view point is shown in Fig. 7. A zoomed-in view of the blurred occluders (in this case people) is shown in Fig. 7(c). The latent image reblurred using the estimated ω and registered with the blurred and occluded observation is shown in Fig. 7(d). The residual image (Fig. 7(e)) reveals that the dark occluders have been accurately detected by the proposed method. Fig. 8 depicts an example from VIRAT dataset in which, after the process of registration, the moving truck has been detected correctly as occlusion in Fig. 8(d).

4   Restoration of foggy and motion-blurred scenes

In a medium such as fog, light rays get attenuated space-variant before they reach the camera sensor. The scattering coefficient of such a medium is high and each ray gets attenuated by a multiplicative factor that is an exponentially decaying function of the scene depth and scattering coefficient. Most of the methods for defogging need more information than just a single image. In [41], multiple images of the same scene under different atmospheric conditions were used. In [42], prior knowledge of the depth map was used while in [43] a polarization filter was employed.
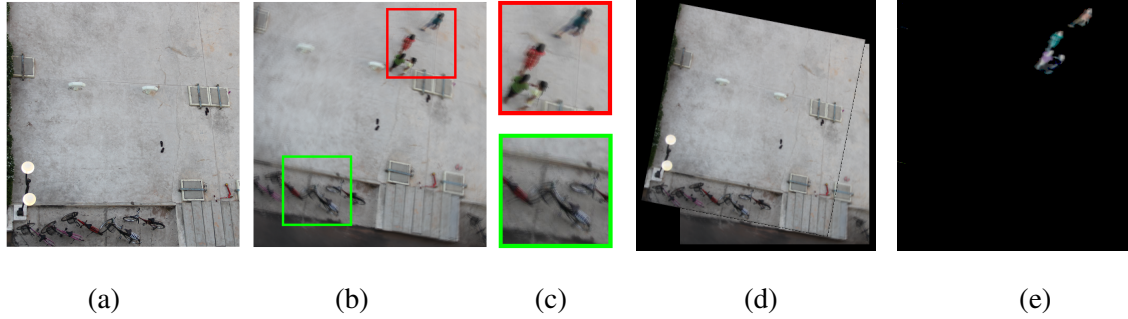
Figure 7: An aerial view of a parking lot. (a) Latent (unblurred) image, (b) blurred and occluded observation taken from a different view point, and (c) zoomed-in regions showing the presence of non-uniform blur, (d) reblurred latent image registered with the blurred and occluded observation, and (e) residual image.
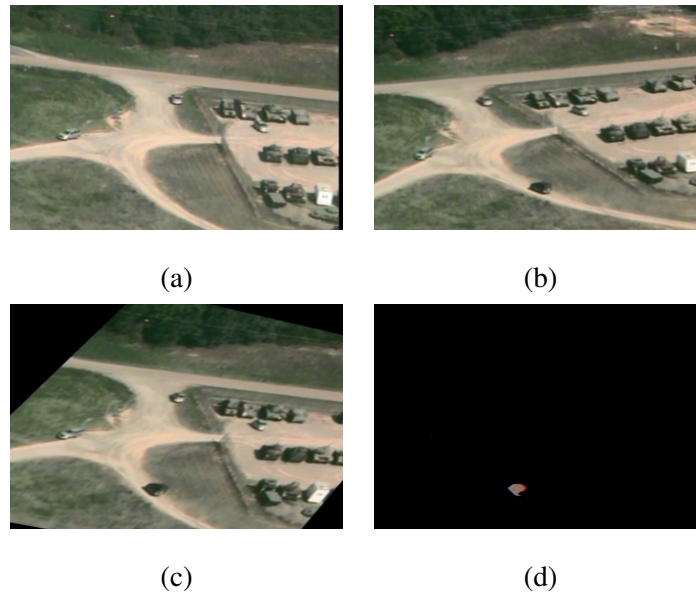

Figure 8: VIRAT dataset. (a) Blurred scene with no activity. (b) Blurred scene with activity. (c) Image of (b) after registration. (d) Detected activity.

Recently, single image defogging techniques have been proposed in [44, 45]. A refined image formation model was used in [44] to account for surface shading while in [45], the dark channel prior was used.

We investigate the problem of motion blur due to camera shake in a foggy scene and mathematically justify that the motion-blurred and medium-attenuated scene radiance can be modelled using a modified image formation model in fog wherein the motion-blurred radiance replaces the original scene radiance. The dark channel prior is then used to get a coarse depth map of the scene along with the blurred scene radiance. For in-plane translational motion-blurred images, the depth map is directly used for restoration by exploiting the scaling relationship among the motion blur kernels at different depths. For the case of general camera motion, blur kernels are estimated at multiple patches and they could be at different depths. The blur kernel at a point is written as a convex sum of weighted impulses located at the displacements undergone by a pixel under the influence of camera motion. These kernels help deduce the camera motion responsible for the observed blurring. While the blur induced by rotation is depth-invariant, the coarse depth map derived from fog is used to scale the translational motion by scene depth. Fog is also used as a cue to segment road scene images into road, left, right and sky planes. Finally, knowledge of the camera motion is used in conjunction with segmentation to deblur each of these planes. The framework can be comfortably applied on aerial images too wherein only the ground plane exists.

4.1   Image formation in haze and blur

In the absence of motion blur, a color image captured in fog can be written [44, 45] as

$$\mathbf{I}(\mathbf{x}) = \mathbf{t}(\mathbf{x})\mathbf{J}(\mathbf{x}) + (1 - \mathbf{t}(\mathbf{x}))\mathbf{A} \tag{20}$$

where $\mathbf{J}$ is the scene radiance, $\mathbf{A}$ (a constant) is the ambient light in the atmosphere, $\mathbf{x}$ is a vector representing the spatial location of a pixel, and $\mathbf{t}$ is related to the medium scattering coefficient (β) and scene depth ($\mathbf{d}$) through the relation $\mathbf{t}(\mathbf{x}) = \exp(-\beta\mathbf{d}(\mathbf{x}))$. If the camera underwent shake during exposure, then the captured image would be a blurred version of the original image. We can write this as

$$\mathbf{B}(\mathbf{x}) = \frac{1}{T_e} \int\limits_0^{T_e} I(\mathbf{x} - \mathbf{x}_\tau) \, \mathrm{d}\tau \tag{21}$$

where T_e is the camera exposure time and $\mathbf{x}_\tau$ is the motion path of pixel $\mathbf{x}$. Therefore, the blurred image captured in fog can be written in integral form as

$$\mathbf{B}(\mathbf{x}) = \frac{1}{T_e} \int_0^{T_e} [J(\mathbf{x} - \mathbf{x}_\tau) \exp(-\beta \mathbf{d}_{\mathbf{x} - \mathbf{x}_\tau})$$

$$+ \mathbf{A}(1 - \exp(-\beta \mathbf{d}_{\mathbf{x} - \mathbf{x}_\tau}))] \, \mathrm{d}\tau$$

Since neighboring points $\mathbf{x}$ and $(\mathbf{x} - \mathbf{x}_\tau)$ will be approximately at the same depth, we can write

$$\mathbf{B}(\mathbf{x}) = \frac{1}{T_e} \exp(-\beta \mathbf{d}_{\mathbf{x} - \mathbf{x}_\tau}) \int_0^{T_e} J(\mathbf{x} - \mathbf{x}_\tau) \, \mathrm{d}\tau$$

$$+ (1 - \exp(-\beta \mathbf{d}_{\mathbf{x} - \mathbf{x}_\tau})) \mathbf{A}$$

It is interesting to observe that the above equation is very similar in form to the fog image formation model (equation (20)) except that $\mathbf{J}$ has been replaced by the term $\frac{1}{T_e} \int_0^{T_e} J(\mathbf{x} - \mathbf{x}_\tau) \, \mathrm{d}\tau$, which is, in fact, the blurred radiance of the scene captured in a lossless medium. Consequently, we can write the modified image formation model taking both blur and fog into account as

$$\mathbf{I}(\mathbf{x}) = \mathbf{t}(\mathbf{x})\mathbf{J_B}(\mathbf{x}) + (1 - \mathbf{t}(\mathbf{x}))\mathbf{A} \tag{22}$$

where $\mathbf{J_B} = \frac{1}{T_e} \int_0^{T_e} J(\mathbf{x} - \mathbf{x}_\tau) \, \mathrm{d}\tau$. The implication of this result is that a coarse depth map and blurred radiance can be obtained using dark channel prior [45] akin to the case of a pure foggy image. We express the dark channel of the motion blurred radiance as

$$\mathbf{J}_{\mathbf{B}dark}(\mathbf{x}) = \min_{\mathbf{s} \in \omega(\mathbf{x})} \min_c (\mathbf{J}_{\mathbf{B}c}(\mathbf{s})) \tag{23}$$

where the color-channel is c and $\omega(\mathbf{x})$ is a small neighborhood around pixel $\mathbf{x}$. As discussed in [45], the principle behind a dark channel prior is that one of the color channels in the proximity of a point is close to zero for outdoor scenes. Since $\mathbf{J_B}$ is, in fact, a blurred version of the scene radiance, it stands to reason that $\mathbf{J}_{\mathbf{B}dark}(\mathbf{x})$ will be close to zero. Consequently, equation (22) yields $\mathbf{t}(\mathbf{x})$ as

$$\mathbf{t}(\mathbf{x}) = 1 - \min_{\mathbf{s} \in \omega(\mathbf{x})} \left(\min_c \left(\frac{\mathbf{I}_c(\mathbf{s})}{\mathbf{A}_c} \right) \right) \tag{24}$$

A scaled depth of the scene can thus be inferred using the relationship between transmission and depth. Since the airlight can be computed easily, we can obtain $\mathbf{J_B}$ from equation (22). Because of the min operation, blocky artifacts can appear in the depth map as well as in the recovered scene radiance. These can, however, be removed by using closed-form matting followed by bilateral filtering.

18

4.2 Motion Deblurring in fog/haze

In this section, we first discuss deblurring under pure translations and then examine the case of general camera motion. The blurred scene radiance and depth map are obtained as discussed in the earlier section.

4.2.1 Translational motion

The motion blur kernel at a point $\mathbf{x}$ can be written as

$$h(\mathbf{x}; \mathbf{s}) = \frac{1}{T_e} \int_0^{T_e} \delta(\mathbf{s} - \bar{\mathbf{x}}(\mathbf{x}, \tau)) \, \mathrm{d}\tau \tag{25}$$

where T_e is the camera exposure time while $\bar{\mathbf{x}}$ is the motion path of pixel $\mathbf{x}$ as a function of τ. For pure camera translations, the blur kernels at different pixels on the image are related by their relative depths. If d_0 is the scene depth of a pixel where the translational motion blur kernel is $h_0(\mathbf{s})$, then the blur kernel at any other point $\mathbf{x}$ that is at a depth $d(\mathbf{x})$ is

$$h(\mathbf{x}; \mathbf{s}) = k^2(\mathbf{x}) h_0(k(\mathbf{x})\mathbf{s}) \tag{26}$$

where $k(\mathbf{x})$ is the ratio $\frac{d(\mathbf{x})}{d_0}$. Using this relationship, we can obtain the blur kernel at any point in the image from a reference kernel and the relative depth map. The blurred radiance can be expressed as

$$\mathbf{J}'_{\mathbf{B}} = \mathbf{C}\mathbf{I}' \tag{27}$$

where the matrix $\mathbf{C}$ consists of blur kernels at every pixel lexicographically ordered in each row, the vector $\mathbf{J}'_{\mathbf{B}}$ is obtained by lexicographically ordering $\mathbf{J}_{\mathbf{B}}$, while the original image is represented as column vector $\mathbf{I}'$. The above equation is solved for $\mathbf{I}'$ using conjugate gradient squared method since the blur matrix $\mathbf{C}$ is asymmetric. A look-up table is constructed for discrete steps of depth values and the matrix multiplication $\mathbf{C}$ is equivalently done with a pixel-wise blur operation. Thus, the original (deblurred) radiance can be recovered.

4.2.2 General camera motion

Deblurring in the presence of general camera motion is more involved since the blur kernels at different locations are no longer related by just a scale factor. In our approach, we first use fog itself as a cue to segment the scene into road, left, right and sky planes using the planar road constraint [46]. Following this, the motion blur induced on each of

these planes is inferred based on the estimated camera motion. Finally, each of these planes is deblurred using the projective motion-blur model. For aerial imagery, only the ground plane exists.

4.2.3 Scene segmentation

For road scenes, under the planar assumption [46], an inverse relationship exists between the image height (i.e., the row number) and the scene depth in the road region. For a $3D$ point (X, Y, Z), the X and Y co-ordinates would be constant for all the points on the road (assuming a flat planar surface for the road). Thus, we can write $\mathbf{t}$ as a function of y as

$$\mathbf{t}(\mathbf{x}) = \exp\left(-\beta\mathbf{d}(\mathbf{x})\right) = \exp\left(-\beta\frac{fY}{y}\right) \tag{28}$$

From equation (20), we get the original scene radiance as

$$\mathbf{J} = \frac{\mathbf{I} - (1 - \mathbf{t})\mathbf{A}}{\mathbf{t}} \tag{29}$$

Thus, note that for image points that do not correspond to the road, $\mathbf{t}$ will be underestimated by equation (28) and this will lead to negative values for $\mathbf{J}$ since $(1 - \mathbf{t})$ will be high. We make all the positive-valued pixels obtained from equation (29) equal to 1 and the rest to 0 to get a binary image. We use median filtering to smoothen this image. The result is a segmentation of the scene into road, left, right and sky planes. Approximate distances to each of these planes are assumed to be known apriori.

(a)   (b)

Figure 9: (a) Foggy road image. (b) Segmentation result.

The segmentation result obtained for the foggy road image of Fig. 9(a) is shown in Fig. 9(b). The white pixels depict the road and sky regions in Fig. 9(b).

As discussed earlier, a planar homography relates any two views of the same planar scene and is given [47, 48] by

$$\mathbf{H} = \mathbf{K}\left(\mathbf{R} + \frac{1}{d}\mathbf{T}\mathbf{n}^T\right)\mathbf{K}^{-1} \tag{30}$$

20

where $\mathbf{K}$ is the camera calibration matrix, $\mathbf{R}$ is the $3D$ rotation matrix, $\mathbf{T}$ is the $3D$ translation vector, while $\mathbf{n}$ and d are the plane normal and perpendicular distance from the camera center, respectively. A motion blurred image of an arbitrary plane can be written as

$$\mathbf{B}(\mathbf{x}) = \sum_\lambda \mathbf{h_T}(\lambda)\mathbf{I}(\mathbf{H}_\lambda \mathbf{x})$$

where $\mathbf{H}_\lambda$ is the homography matrix that relates the view λ to the reference view and $\mathbf{h_T}(\lambda)$ is the fraction of time that the camera spends in view λ during the exposure time.

The cross-section of a plane at a particular depth $Z = c$ can be written as

$$n_1 X + n_2 Y + n_3 c = d$$

The projection of any point $\left(a, \frac{d - n_1 a - n_3 c}{n_2}, c\right)$ on this line onto the image plane would be $x = \frac{a}{c}$ and $y = \frac{d - n_1 a - n_3 c}{n_2 c}$. For pure in-plane translation and rotation of the camera, assuming the focal length to be unity (for simplicity), the plane image point would move to the position given by a $2D$ transformation

$$\mathbf{Hx} = \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) \\ \sin(\theta_z) & \cos(\theta_z) \end{bmatrix} \begin{bmatrix} \frac{a}{c} \\ \frac{d - n_1 a - n_3 c}{n_2 c} \end{bmatrix} + \begin{bmatrix} \frac{T_x}{c} \\ \frac{T_y}{c} \end{bmatrix} \tag{31}$$

The displacement of a small patch lying on the plane would be given by the above relation (31).

The point spread function at a point $\mathbf{x}$ can be related to the camera motion ($\mathbf{h_T}(\lambda)$) as

$$h(\mathbf{x}; \mathbf{s}) = \int_\lambda \mathbf{h_T}(\lambda)\delta(\mathbf{s} - (\mathbf{x} - \mathbf{H}_\lambda \mathbf{x}))\, \mathrm{d}\lambda$$

Using this relationship, we can write a matrix equation that relates blur kernels to camera motion as

$$\mathbf{h} = \mathbf{M}\mathbf{h_T} \tag{32}$$

where $\mathbf{h}$ is a column vector of blur kernels lexicographically ordered and $\mathbf{M}$ is a matrix that holds the value 1 at rows corresponding to the displacement of a patch under camera transformation $\mathbf{H}_\lambda$. The depth $\mathbf{d}$ is factored into the computations using the coarse depth map estimated from equation (24). Following [9], we consider only the three parameters (t_x, t_y and θ_z) when estimating camera motion. Equation (32) is solved with $L1$ norm sparsity constraint on $\mathbf{h_T}$.

| (a) | (b) | (c) |

Figure 10: (a) Foggy and motion-blurred road image. (b) Defogged image. (c) Defogged and deblurred result.

4.2.4 Motion deblurring of planes

The projective motion Richardson-Lucy algorithm of Tai et al. [6] deblurs the image of a fronto-parallel scene under known camera motion. We can deblur an arbitrary plane after estimating the camera motion using the technique in the previous section. The likelihood probability $P(\mathbf{B}, \mathbf{h_T}|\mathbf{I})$ is modeled using Poisson distribution and the Kuhn-Tucker conditions are used to arrive at the iterative update equation for the original radiance given the camera motion $\mathbf{h_T}$ as

$$\mathbf{I^{n+1}} = \mathbf{I^n} \sum_\lambda \mathbf{h_T}(\lambda)\mathbf{E^n}(\mathbf{H}_\lambda^{-1}\mathbf{x}) \tag{33}$$

where $\mathbf{E}^n$ is an error matrix given by $\mathbf{E}^n = \frac{\mathbf{B}}{\mathbf{B}'^n}$ where the division is point-wise. Note that $\mathbf{B}'^n$ is the blurred scene radiance of the plane under consideration and can be obtained from $\mathbf{I^n}$ as

$$\mathbf{B}'^n = \sum_\lambda \mathbf{h_T}(\lambda)\mathbf{I}^n(H_\lambda\mathbf{x}) \tag{34}$$

In addition, the total variation prior [6] is used to preserve edges.

4.3 Experimental results

In this section, we give representative results for restoring foggy and motion blurred images in real scenarios.

We first examine a road scene image that had incidental motion blur which was non-uniform. This image was taken outside our department with a mobile phone camera. Since it was raining, the collective appearance of raindrops made the scene appear foggy (Fig. 10(a)). The blurred scene radiance and coarse depth map were obtained from equation (22). We computed blur kernels at four different locations in the image. Using the depth map and these blur kernels, the camera motion was estimated using equation (32). The foggy images were segmented into road, left, right and sky

planes using the planar road constraint. Each of these planes was then deblurred using equation (33). The defogged and deblurred result is shown in Fig. 10(c) wherein we can observe the improvement that accrues due to deblurring.

The first row of Fig. 11 depicts deblurring as well as dehazing of a real aerial image from VIRAT dataset. Note the improvements in the restored image over the observation. We also show results on another real aerial image (second row of Fig. 11) kindly provided by Dr. Steve Sudarth from Transparent Sky, USA. The appearance after deblurring and dehazing is quite striking.

(a)   (b)

Figure 11: VIRAT dataset (first row) and another real aerial image (second row). (a) Hazy and motion-blurred image. (b) Dehazed and deblurred result.

5 Multi-frame blind super-resolution of non-uniformly blurred images

The effective resolution of an imaging system is limited not only by the physical resolution of its image sensor but also by blur. If the blur is present, super-resolution makes little sense without removing the blur. The presence of a spatially

23

varying blur makes the problem much more challenging and for the present, there are almost no algorithms designed specifically for this case. The critical part of such algorithms is precise estimation of the varying blur, which depends to a large extent on a specific application and type of blur. However, non-uniform deblurring and super-resolution, though two extensively studied topics, have mostly been dealt with independently. The importance of deriving a high-resolution image from low-resolution observations is very well-known and is an active area of research. However, what is not obvious is the fact that a motion blurred image can be used to perform SR. This is due to the fact that the motion blurred image of a planar scene stems from the weighted average of a sequence of geometric warps of the original scene.

A large number of papers address the standard SR problem when the images are not blurred. A good survey can be found for example in [49] and [50]. Maximum likelihood, maximum a posteriori (MAP), the set thoeretic approach using projection on convex sets, and fast Fourier techniques can all provide a solution to the SR problem. Spatial-domain SR approaches are popular as they can accommodate complex priors and can handle even non-global motion [49]. A multi-frame SR method to detect and reconstruct small rigid moving objects with translucent pixels is elaborated in [51]. Movements between adjacent frames are generally assumed to be smooth [52], but due to object and/or camera motion, the shifts in the frames can be space-variant. Brox and Malik [53] employ an optical flow method which is well-suited for large local motion. However, state-of-the-art SR techniques achieve remarkable results of resolution enhancement only in the case of no blur.

Sroubek and Cristbal [54] propose a unifying method that simultaneously estimates the camera motion and the HR image from a set of blurred LR observations without any prior knowledge of the blurs and the original image. Harmeling et al. [55] solve the same multi-frame super-resolution problem using an incremental expectation maximization (EM) framework that does not require explicit image or blur priors. But these approaches are based on the standard convolution model and is restricted to translational motion. They do not tackle the more real case of space-varying blur. A naive approach to tackling this scenario includes applying super-resolution using the convolution model to small space-invariant regions in the image and then sewing up the patches. Unfortunately, it is not easy to sew the patches together without artifacts on the seams. An alternative way is first to use the estimated PSFs to approximate the spatially varying PSF by interpolation of adjacent kernels and then compute the image of improved resolution. The main problem of these naive procedures is that they are relatively slow, especially if applied on too many positions. We present a new joint approach to the super-resolution and non-uniform deblurring problem. We use HR PSFs estimated at a few locations in the LR frames to compute the HR TSFs which reveal the camera motion

during exposure for the latent HR image. The restoration step is finally carried out using suitable regularization terms on the image.

5.1 The SR motion blur model

In the discrete domain, the operation of blurring and downsampling can be represented by the following equation.

$$g\left(i,j\right) = D\left(\sum_{l=1}^{N_T} \omega\left(l\right) f\left(\mathcal{H}_l\left(i,j\right)\right)\right) \tag{35}$$

Here $f(i,j)$ denotes the latent HR image of the scene. $g(i,j)$ is the blurred LR image. $\mathcal{H}_l(i,j)$ denotes the image coordinates when a homography $\mathcal{H}_l$ is applied on the point (i,j). (When $\mathcal{H}_l(i,j)$ takes non-integer values, we assign values to the pixels neighboring $\mathcal{H}_l(i,j)$ by bilinear interpolation principle.) D is the downsampling operator or the decimation operator that models the function of the CCD sensors. A multi-frame method that uses the information present in adjacent frames is used to increase the spatial resolution of the super-resolved image. If we have K LR frames, then we can rewrite equation (35) as follows

$$g_k\left(i,j\right) = D\left(\sum_{l=1}^{N_T} \omega_k\left(l\right) f\left(\mathcal{H}_{l_k}\left(i,j\right)\right)\right) \text{ where } k = 1, 2, ..., K. \tag{36}$$

where $g_k\left(i,j\right)$ is the kth LR observation and ω_k and $\mathcal{H}_{l_k}$ are the associated TSFs and homographies.

5.2 The proposed method

Consider K blurred LR observations $g_1, g_2, ..., g_K$ of a scene which are related to the latent HR image f through the HR TSFs $\omega_1, \omega_2, ..., \omega_K$. The objective is to estimate f from $g_1, g_2, ..., g_K$. To compute the HR TSFs, we re-formulate the TSF estimation technique in [56] to the SR scenario by using HR blur kernels estimated at a few points in the LR frames. With the knowledge of the HR TSFs, we solve for the latent HR image f within a regularization framework. The details of these steps are explained in the following subsections.

5.2.1 HR PSF estimation

Our first step is to estimate HR blur kernels at different locations in the LR frames. To this end, we use [57] to determine regions in the image with good texture and long edges that are suitable for blur kernel estimation. From this subset, we select N_p point locations, and blurred image patches from the K LR observations (denoted by $g_1^1 g_1^2 ... g_1^{N_p}$,

25

$g_2^1 g_2^2 ... g_2^{N_p}, ..., g_K^1 g_K^2 ... g_K^{N_p}$) are cropped around these points. Within each set of patches $(g_1^i, g_2^i, ..., g_K^i)$, we assume the blur to be space-invariant and use the blind SR technique in [54] to get the HR blur kernels $h_1^i, h_2^i, ..., h_K^i$. The method in [54] requires that the number of observations K needed to estimate the HR blur kernels should be greater than the SR factor (ϵ) squared i.e., $K > \epsilon^2$. We found in our experiments that the estimates of the blur kernels from the method in [54] are quite accurate.

5.2.2 HR TSF from HR PSFs

Our next objective is to estimate the HR TSF ω_k that concurs with the N_p observed HR blur kernels $h_k^1 h_k^2 ... h_k^{N_p}$ for $k = 1, 2, ...K$. Note that the TSF $\omega_k \in R^{N_T}$ will be a sparse vector in practice because the camera motion during exposure would result in very few transformations out of all the possible elements of T. Hence while solving for ω_k, we impose a sparsity constraint for regularization. Following [56], we express the blur kernel h_k^i as $h_k^i = M_k^i \omega_k$ for $i = 1, 2, ...N_p$ since each component of the blur kernel h_k^i at a given pixel is a weighted sum of the components of the TSF ω_k. Here, M_k^i is a matrix whose entries are determined by the location of the blur kernel and the bilinear interpolation coefficients. Note that the N_p point locations were chosen on the LR grid and patches were cropped around these points from the LR frames. Since the TSFs are being estimated on an HR grid, the N_p point locations should be scaled by the SR factor ϵ, and our TSF estimation step differs from the method proposed in [56] in this important respect. If the number of elements in the blur kernel is N_h, then the size of the matrix M_k^i will be $N_h \times N_T$. By stacking all the N_p blur kernels as a vector $\overline{h_k}$, and suitably concatenating the matrices M_k^i for $i = 1, 2, ..., N_p$, the HR blur kernels can be related to the HR TSF as

$$\overline{h_k} = M_k \omega_k \tag{37}$$

The matrix M_k is of size $N_p N_h \times N_T$. To get an estimate of the HR TSF that is consistent with the observed HR blur kernels and is sparse as well, we minimize the following cost

$$\underset{\omega_k}{\operatorname{argmin}} \left\{ ||\overline{h_k} - M_k \omega_k||_2^2 + \lambda ||\omega_k||_1 \right\} \tag{38}$$

where the positive scalar λ controls the extent of sparsity of the estimated HR TSF. We estimate each HR TSF $\omega_1, \omega_2, ..., \omega_K$ separately by minimizing the above cost function using the toolbox in [29].

5.2.3 Deblurring

Equation (36) can be expressed in the matrix-vector notation [54] as

$$g_k = DH_k f \tag{39}$$

where H_k and D are the matrices that perform the non-uniform blurring and downsampling operations respectively.

Once $\omega_1, \omega_2, ..., \omega_K$ are known, we formulate the energy function based on the observation error and a regularization term as

$$E(f) = \sum_{k=1}^{K} ||DH_k f - g_k||_2^2 + \alpha f^T L f \tag{40}$$

where L is the discrete form of the variational prior and is a positive semidefinite block tridiagonal matrix [54] constructed of values depending on the gradient of f. The rationale behind the choice of this prior is to constrain the local spatial behaviour of the images. While in smooth areas, it has the same isotropic behaviour as the Laplacian, it also preserves edges. The disadvantage is that it is highly non-linear and the half-quadratic algorithm has to be used to minimize (40). α represents the weight of the image smoothness term.

$$\frac{\partial E}{\partial f} = 0 \Longleftrightarrow \left(\sum_{k=1}^{K} H_k^T D^T D H_k + \alpha L \right) f = \sum_{k=1}^{K} H_k^T D^T g_k \tag{41}$$

H^T is equivalent to blurring through a TSF except for the fact the warping to be applied is $\mathcal{H}_l^{-1}$ instead of $\mathcal{H}_l$ [56]. The matrix D^T spreads equally the intensity in LR to ϵ^2 pixels in HR.

We use the method of conjugate gradients to solve (41) and then adjust the solution f to contain values in the admissible range, typically, the range of values of g_k

5.3 Experimental results

We begin with a synthetic example. The image of a tablecloth (of size 530×640 pixels) shown in Fig. 12(a) was used as the latent HR image of the scene. To simulate the shake incurred by a camera during exposure, we manually generated camera motion with a connected path in the motion space and initialized the weights. Five different HR TSFs were synthesized thus and the five corresponding blurred LR frames were generated from the latent image by first blurring (using the TSF model) and then downsampling by a factor of two. For an SR factor of 2, the method in [54] needs a minimum of five LR images for the computation of the HR PSF. The parameters of the TSF ranged as follows: θ ranged between -2 to 2 degrees in steps of 0.5, t_x and t_y ranged between -3 to 3 pixels in steps of one pixel.
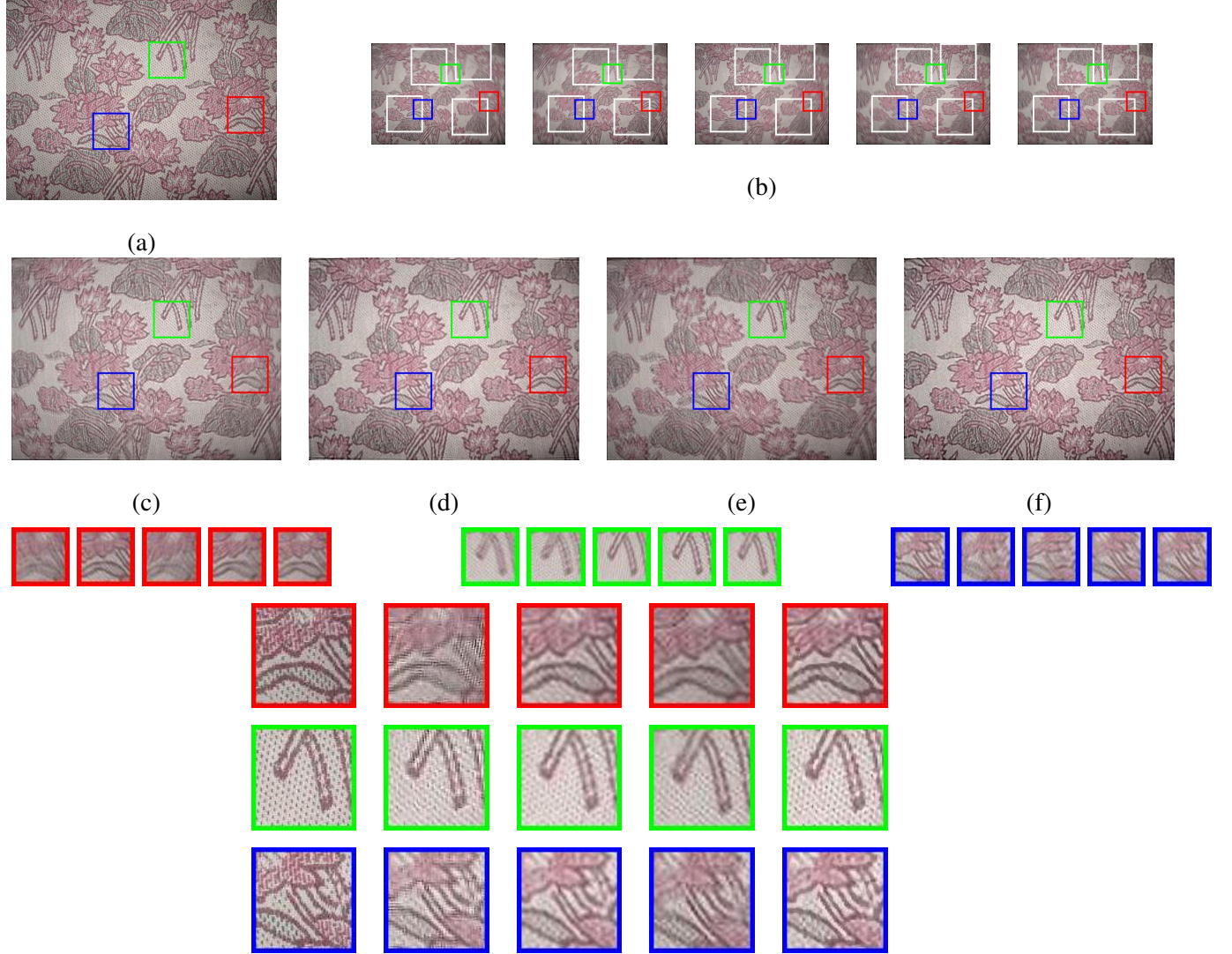
Figure 12: (a) Latent HR image, (b) blurred LR observations, (c) Sroubek *et al.* [54], (d) upsampled output of Paramanand *et al.* [56], (e) upsampled output of Whyte et al. [10], and (f) our output. Row 3: Zoomed-in regions from the blurred LR observations. Rows 4,5,6: Zoomed-in regions from the original image (first column), Sroubek *et al.* [54] (second column), upsampled output of Paramanand *et al.* [56] (third column), upsampled output of Whyte et al. [10] (fourth column) and our output (fifth column).
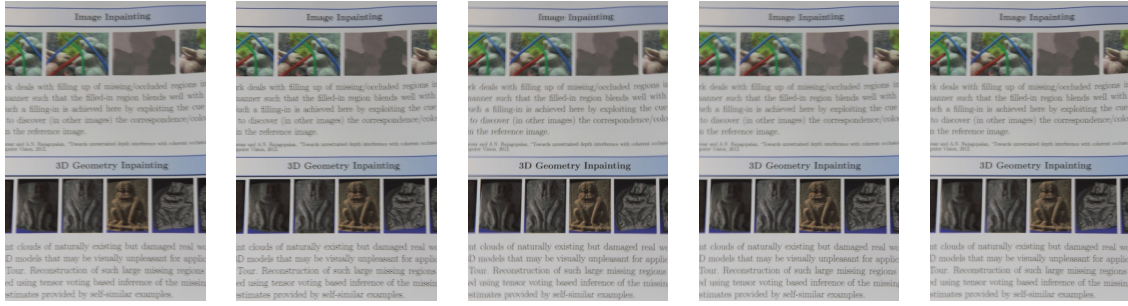
(a)
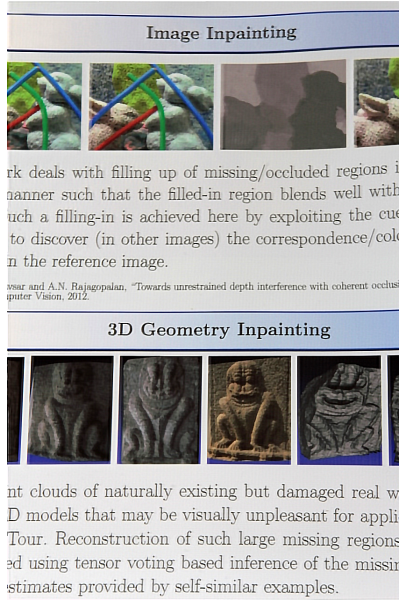
(b)

(c)

(d)

Figure 13: Real experiment 1: (a) Blurred LR observations, (b) our output, (c), (d) zoomed-in regions from (a) and (b).

We selected four spatially separated image patches having good texture from the 265×320 LR observations and used the algorithm in [54] to determine the HR blur kernels corresponding to these patches. The blurred LR observations with the selected patches (enclosed in white boxes) are shown in Fig. 12(b). The HR TSFs were then computed from these HR blur kernels using the method described in section 5.2. The deblurred HR image is shown in Fig. 12(f). The output obtained by using the convolution model in [54] is shown in Fig. 12(c). For comparison, we also upsampled by a factor two the LR output image of [56] using bicubic interpolation. This is shown in Fig.12(d). It is to be noted that all five LR observations were given as input to the algorithm in [56]. Another comparision, obtained by first deblurring one of the LR frames using the technique in [10] and then upsampling the result, is provided in Fig. 12(e). In row 3 of Fig. 12, we show three sets of zoomed-in regions from the five blurred LR observations. Zoomed-in patches from the original image (first column), output of [54] (second column), upsampled output of [56] (third column), upsampled output of [10] (fourth column) and our output (fifth column) are shown in rows 4,5 and 6. We observe that our output is sharper and compares more closely to the original than all the other methods. Fig. 13 depicts super-resolution results for a real case. Note the improvement in readability of text that accrues after super-resolution.

6   Conclusions, highlights and future directions

In this report, we investigated the problem of processing images and videos in the presence of spatially-varying motion blur. Intermediate results were produced regularly and information exchanged between PI and Dr. Guna from AFRL to facilitate interaction and forward movement.

Key accomplishments

Analytical expressions were derived to relate the degree of blur to the nature of the scene and camera motion, and robust methods were proposed for image deblurring.

- *Based on the notion of a global transformation spread function, we propose a formulation that can effectively restore space-variant motion blurred images affected by arbitrarily-shaped blur kernels.*

Our efforts will be valuable in mitigating the effects of motion blur in practical situations.

Following this, we addressed the related problem of automatically detecting occluded regions given a pair of images of a scene taken from different viewpoints. The occlusion could be due to single or multiple objects. What makes this problem difficult for surveillance applications is the fact that the image pair is both geometrically and

30

photometrically out-of-sync.

- *We have presented a unified framework based on sparsity prior for automatically detecting occluder(s). The method is reasonably robust to non-uniform motion blur as well as variations in camera pose (without the need for deblurring).*

Given the distances involved in aerial imagery and the fact that it is not uncommon to use plexiglass for lens protection, the problems of deblurring and dehazing often co-occur.

- *We have expanded the scope of the restoration problem to not only include deblurring but also dehazing with the aim of improving loss of visibility due to poor contrast.*

Finally, we addressed the problem of estimating the latent high-resolution (HR) image from a set of non-uniformly blurred low-resolution (LR) images.

- *We have presented a new framework that judiciously exploits sub-pixel motion arising from motion blur to perform super-resolution of images.*

The HR TSF, which reveals the camera motion during exposure for the latent HR image, is computed from the LR frames.

Each of the above-mentioned efforts was comprehensively analyzed and validated with synthetic as well as real data.

- *While some of our new findings have already been published in peer-reviewed avenues, some more are under review in prestigious international journals.*

Expected impact

The related problems of image restoration, registration, dehazing, and superresolution, all in the presence of blurring due to camera motion, that have been investigated within the ambit of this proposal stand at the cutting edge of research. The works undertaken in this proposal will go a long way in providing much-needed theoretical insights and a strong mathematical underpinning into the relatively less-understood but all-prevalent motion blur. The potential ramifications of the work undertaken are many, some of which are discussed next.

A sharp image is beneficial not only from the perspective of visual appeal but also because it forms the basis for moving object tracking, change detection, robust feature extraction etc. There is another novel spin to this problem.

Because the motion blur results from platform motion, knowledge of motion blur can prove to be a valuable cue in stabilizing the captured video. Our work on registration can form the basis for handling view-point changes as these can be captured through poses of the camera as part of motion blur estimation and the intensities can be compared through weighted geometric warps of the images. The twin problems of deblurring and dehazing open up exciting avenues for future research. If this capability can be achieved on board and in real-time, it can be very valuable for applications involving aerial surveillance. Our work on superresolution will be especially valuable while attempting to track targets that occupy just a few pixels. A possible extension to our work will be to allow for depth variations in the scene.

Future directions

The theories proposed within the ambit of this proposal are not limited to aerial images and can potentially be used in any situation in which the scene can be modeled as approximately planar. This could include even human faces, for example. The importance of face recognition cannot be underestimated for homeland security. Current works on face recognition assume that the face is reasonably focused or at worst suffers from uniform motion blur. Our work can help generalize the theory of face recognition to non-uniform blurring situations too. The future of face recognition research is moving towards recognition of subjects in motion and our work will be critical to this futuristic scenario.

Although our focus is on camera motion, the entire theory is also applicable to the case when there is a single moving object relative to the camera. An interesting off-shoot to this problem arises when you have moving camera as well as moving targets. What is fascinating is that the motion blur can potentially be used for segmenting independently moving objects in a scene. This is a very fertile area and there is a lot of excitement surrounding it. We believe that the natural connect that exists between motion blur and alpha-mattes should be exploited to segment dynamic scenes. As a related spin-off to the motion deblurring problem, it should be possible to exploit the proposed framework to fork into the exciting area of image forensics to even detect splicing in images.

We shall continue our investigations into the problems of image registration, occlusion detection, and super-resolution, especially from the viewpoint of low-rank, sparse error matrix decomposition. It is envisioned that this will lead to a robust analytical and computational framework that can be exploited to address practical scenarios. In addition, we shall also focus on the recovery of 3D information of scenes imaged by a moving camera. Here, the motion blur will be harnessed as a cue for depth since the extent of motion blur at an image point is dictated both by scene structure and camera motion. We shall also investigate the related spin-off problems of splicing detection in the

exciting domain of image forensics.

Summary statement

The works carried out under this proposal can help foster excellence in basic research, create new scientific understanding, and make available unforeseen and innovative technological options for the scientific community. They can revolutionalize and profoundly impact the future capabilities of the AFOSR/AFRL in their ability to harness valuable information from captured image data giving them a distinct technological edge in the air to meet the growing challenges of the future.

7 Research publications arising out of the efforts

1. P. Abhijith, A.N. Rajagopalan and Guna Seetharaman. Registration and occlusion detection in motion blur, in *Proc. IEEE International Conference on Image Processing* (ICIP) Melbourne, Sep. 2013.

2. T. Veeramani, A. N. Rajagopalan and Guna Seetharaman. Restoration of foggy and motion-blurred road scenes, in *Proc. IEEE International Conference on Image Processing* (ICIP) Melbourne, Sep. 2013.

3. P. Rao, A.N. Rajagopalan and Guna Seetharaman. Harnessing motion blur to unveil splicing. *IEEE Transactions on Information Forensics and Security* (under review, revised version submitted).

4. T. Veeramani, A. N. Rajagopalan and Guna Seetharaman. Restoration of foggy and motion blurred images. *IEEE Transactions on Geoscience and Remote Sensing* (to be submitted).

References

[1] P. Favaro and S. Soatto, *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion Blur*, Springer-Verlag New York Inc, 2007. 1, 2

[2] D. Doeffinger, *Creative Shutter Speed: Master your Camera's Most Powerful Control*, Wiley Publishing, 2009. 1

[3] T. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, Society for Industrial and Applied Mathematics, 2005. 2, 2

[4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 787–794, 2006. 2, 2.2

[5] Anil K Jain, *Fundamentals of Digital Image Pprocessing*, Prentice-Hall, 1989. 2

[6] Y. Tai, P. Tan, and M. S. Brown, "Richardson-lucy deblurring for scenes under projective motion path," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1603–1618, 2011. 2.1, 2.2, 4.2.4, 4.2.4

[7] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2.1, 2.2

[8] C. Paramanand and A. N. Rajagopalan, "Inferring image transformation and structure from motion-blurred images," in *Proc. British Machine Vision Conference*, 2010. 2.1, 2.2

[9] A. Gupta, N. Joshi, L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *Proc. European Conference on Computer Vision*, 2010. 2.1, 2.2, 4.2.3

[10] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 168–186, 2012. 2.1, 2.2, 12, 5.3

[11] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 874–883, 2005. 2.2

[12] L. Yuan, J. Sun, L. Quan, and H. Shum, "Image deblurring with blurred/noisy image pairs," in *ACM Transactions on Graphics*, 2007, vol. 26, p. 1. 2.2

[13] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2354–2367, 2011. 2.2

[14] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 787–794, 2008. 2.2, 2.3.1, 2.3.2

[15] J. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 567–572, 2012. 2.2

[16] F. Sroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1687–1700, 2012. 2.2

[17] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 2.2

[18] H. Ji and K. Wang, "Robust image deblurring with an inaccurate blur kernel," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1624–1634, 2012. 2.2

[19] H. Ji and K. Wang, "A two-stage approach to blind spatially-varying motion deblurring," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2.2

[20] M Sorel and F Sroubek, "Space-variant deblurring using one blurred and one underexposed image," in *Proc. IEEE International Conference on Image Processing*, 2009, pp. 157–160. 2.2

[21] R. Vio, J. Nagy, and W. Wamsteker, "Multiple-image deblurring with spatially-variant point spread functions," *Astronomy and Astrophysics*, vol. 434, no. 2, pp. 795–800, 2005. 2.2

[22] N. Joshi, S. B. Kang, L. Zitnick, and R. Szeliski, "Image deblurring using inertial measurement sensors," in *Proc. SIGGRAPH*, 2010. 2.2

[23] Y. Tai, N. Kong, S. Lin, and S. Y. Shin, "Coded exposure imaging for projective motion deblurring," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2.2

[24] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Scholkopf, "Fast removal of non-uniform camera shake," in *Proc. International Conference on Computer Vision*, 2011. 2.2

[25] M. Hirsch, S. Sra, B. Scholkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2.2

[26] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2011. 2.2

[27] S. Cho, H. Cho, Y. Tai, and S. Lee, "Non-uniform motion deblurring for camera shakes using image registration," in *SIGGRAPH Talks*, 2011. 2.2

[28] Z. Hu and M. Yang, "Fast non-uniform deblurring using constrained camera pose subspace," in *Proc. British Machine Vision Conference*, 2012. 2.2, 2.3.1

[29] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009. 2.3.1, 5.2.2

[30] R.S. Feris, B. Siddiquie, J. Petterson, Yun Zhai, A. Datta, L.M. Brown, and S. Pankanti, "Large-scale vehicle detection, indexing, and search in urban surveillance videos," *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 28 –42, Feb. 2012. 3

[31] V. Ablavsky and S. Sclaroff, "Layered graphical models for tracking partially occluded objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 9, pp. 1758 –1775, sept. 2011. 3

[32] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, Sept. 2004. 3

[33] K. Shafique and M. Shah, "A non-iterative greedy algorithm for multi-frame point correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 51–65, 2003. 3

[34] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision-Volume 2*, 1999. 3

[35] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of 7th International Joint Conference on Artificial Intelligence - Volume 2*, 1981, pp. 674–679. 3

[36] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007. 3

[37] M. McGuire and H. S. Stone, "Techniques for multiresolution image registration in the presence of occlusions.," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1476–1479, 2000. 3

[38] L. Lelégard, E. Delaygue, M. Brédif, and B. Vallet, "Detecting and correcting motion blur from images shot with channel-dependent exposure time," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-3*, 2012, pp. 341–346. 3

[39] L. Yuan, J. Sun, L. Quan, and H.-Y Shum, "Blurred/non-blurred image alignment using sparseness prior," in *Proceedings of International Conference on Computer Vision*, 2007, pp. 1–8. 3

[40] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009. 3.1

[41] S.K. Nayar and S.G. Narasimhan, "Vision in bad weather," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 820 –827. 4

[42] J.P. Oakley and B.L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 167 –179, Feb 1998. 4

[43] Y.Y. Schechner, S.G. Narasimhan, and S.K. Nayar, "Instant dehazing of images using polarization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Dec 2001, vol. I, pp. 325–332. 4

[44] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 72:1–72:9, Aug. 2008. 4, 4.1

[45] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341 –2353, Dec. 2011. 4, 4.1, 4.1, 4.1

[46] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6 –20, Summer 2012. 4.2.2, 4.2.3

[47] Guna S. Seetharaman, "Handbook of pattern recognition and image processing (vol. 2)," chapter Image sequence analysis for three-dimensional perception of dynamic scenes, pp. 361–403. Academic Press, Inc., Orlando, FL, USA, 1994. 4.2.3

[48] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*, SpringerVerlag, 2003. 4.2.3

[49] S.C Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21–36, 2003. 5

[50] M. Elad S. Farsiu, D. Robinson, , and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, pp. 47–57, 2004. 5

[51] A. W. M. van Eekeren, K. Schutte, and L.J. van Vliet, "Multiframe super-resolution reconstruction of small moving objects," *IEEE Trans. Img. Proc.*, vol. 19, no. 11, pp. 2901–2912, Nov. 2010. 5

[52] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *Trans. Img. Proc.*, vol. 16, no. 2, pp. 479–490, Feb. 2007. 5

[53] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011. 5

[54] F. Sroubek, G. Cristobal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," *Trans. Img. Proc.*, vol. 16, no. 9, pp. 2322–2332, Sept. 2007. 5, 5.2.1, 5.2.3, 5.2.3, 5.3, 12

[55] S. Harmeling, S. Sra, M. Hirsch, and B. Scholkopf, "Multiframe blind deconvolution, super-resolution, and saturation correction via incremental em," in *In Proc. ICIP*, 2010. 5

[56] C Paramanand and A. N Rajagopalan, "Non-uniform motion deblurring for bilayer scenes," in *In Proc. CVPR*, 2013. 5.2, 5.2.2, 5.2.3, 12, 5.3

[57] Z. Hu and M. Yang, "Good regions to deblur," in *Proceedings of the 12th European conference on Computer Vision - Volume Part V*, Berlin, Heidelberg, 2012, ECCV'12, pp. 59–72, Springer-Verlag. 5.2.1